**هوش مصنوعی** **تمرین شماره‌ی دوم**

مدرس: **دکتر فدایی، دکتر یعقوب‌زاده** طراح: محمد امانلو، امین یوسفی

مهلت تحویل: دوشنبه 28 آبان ماه، ساعت 23:59

# Markov Decision Process

## سوال اول) سوالات مفهومی

الف) در کلاس یاد گرفتیم که معادلات بلمن می‌توانند برای توصیف بهره‌وری بهینه در MDPها استفاده شوند. به‌عنوان مرجع، این معادله به این صورت بیان می‌شود:

$$V^*(s) = \max_a \sum_{s'} T(s, a, s')[R(s, a, s') + \gamma V^*(s')]$$

در این معادله، γ چه نامیده می‌شود؟ چرا ضروری است؟ وقتی γ بزرگ‌تر می‌شود چه اتفاقی می‌افتد؟ و اگر کوچک‌تر شود چه تأثیری دارد؟

ب) تفاوت‌های کلیدی بین الگوریتم‌های ارزش‌یابی تکراری و سیاست‌گذاری تکراری (value iteration و policy iteration) چیست و در چه شرایطی ممکن است یکی را بر دیگری ترجیح دهیم؟

ج) سیاست‌گذاری تکراری کی به پایان می‌رسد؟ بلافاصله پس از پایان (بدون محاسبات اضافی) آیا مقادیر سیاست بهینه را داریم؟

د) اگر در طی سیاست‌گذاری تکراری، فقط یک تکرار از به‌روزرسانی بلمن را به جای اجرای کامل تا همگرایی اجرا کنیم، چه تغییری رخ می‌دهد؟ آیا همچنان به سیاست بهینه می‌رسیم؟

## سوال دوم) مسابقه

یک مثال تغییر یافته از مسابقه‌ی ربات خودرو را که در کلاس دیدیم در نظر بگیرید. در این بازی، خودرو به طور تصادفی تعدادی از فضاها را حرکت می‌کند که به‌طور مساوی احتمال دارد ۲، ۳ یا ۴ باشد. خودرو می‌تواند حرکت کند یا متوقف شود اگر مجموع فضاهای حرکت کرده کمتر از ۶ باشد.
اگر مجموع فضاهای حرکت کرده برابر یا بیشتر از ۶ باشد، بازی با پاداش ۰ به پایان می‌رسد. هنگامی که خودرو متوقف می‌شود، پاداش برابر با مجموع فضاهای حرکت کرده (تا حداکثر ۵) خواهد بود و بازی به پایان می‌رسد. برای عمل حرکت پاداشی وجود ندارد.
این مسئله را به‌عنوان یک MDP با وضعیت‌های {0, 2, 3, 4, 5, Done} فرمول‌بندی می‌کنیم.

(الف) تابع انتقال (transition function) برای این MDP چیست؟

(ب) تابع پاداش برای این MDP چیست؟

(ج) ارزش‌یابی تکراری (value iteration) برای ۴ تکرار با ۱ = γ را اجرا کنید.

(د) سیاست بهینه چیست؟

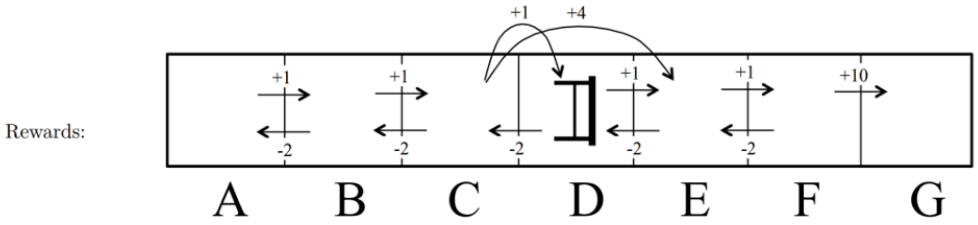
(ه) نتایج چگونه با 0.1 = γ تغییر می‌کند؟ دلیل آن را توضیح دهید.

(و) برای این MDP، دو iteration از تکرار سیاست (policy iteration) را برای یک step از این MDP اجرا کنید، با شروع از سیاست اولیه زیر و با استفاده از مقدار اولیه ۱ = γ. ($\pi_2$ را به همراه مراحل رسیدن به آن، تعیین کنید.)

$\pi_0 = Move, Stop, Move, Stop, Move$

## سوال سوم) دوی با مانع

در نظر بگیرید که یک MDP داریم که یک مسیر دویدن از روی موانع را مطابق شکل زیر نشان می‌دهد. یک مانع در مربع D و وضعیت پایانی در مربع G وجود دارد. عامل می‌تواند به سمت چپ یا راست بدود. اگر عامل در مربع C باشد، می‌تواند به سمت راست بدود ولی به جای آن می‌تواند بپرد، که این عمل ممکن است منجر به سقوط به مربع مانع D شود. پاداش‌ها در زیر نمایش داده شده‌اند و ضریب تخفیف را با مقدار 1 = γ فرض کنید.




اکشن‌ها:

- راست: به طور قطعی به راست حرکت می‌کند. (در خانه C قابل اتخاذ نیست.)
- چپ: به طور قطعی به چپ حرکت می‌کند.
- پرش: به طور تصادفی به راست می‌پرد و فقط برای خانه C قابل اتخاذ است. احتمال موفقیت پرش برابر با 50٪ است.

الف) برای سیاست π که همیشه حرکت مستقیم را پیشنهاد می‌دهد (همیشه راست یا پرش)، مقدار $V^{\pi}(C)$ را محاسبه کنید.

ب) دو بار پیمایش ارزش (value iteration) را انجام دهید و مقادیر زیر را حساب کنید. مقداردهی اولیه همه ارزش‌ها برابر صفر است.

- $V_2(B)$
- $Q_2(B, Right)$
- $Q_2(B, Left)$

ج) برای خانه‌های خالی جدول زیر، مقادیر Q-valueها را با بروزرسانی‌هایی که از اعمال انتقال مشخص شده برای Q-learning به دست می‌آیند، پر کنید. از نرخ یادگیری α = 0.5 استفاده کنید و فرض کنید همه Q-valueها در ابتدا برابر صفر بودند. خانه‌هایی که تغییری نمی‌کنند خالی بگذارید.

Episode

| $s$ | $a$ | $r$ | $s$ | $a$ | $r$ | $s$ | $a$ | $r$ | $s$ | $a$ | $r$ | $s$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C | *jump* | +4 | E | *right* | +1 | F | *left* | -2 | E | *right* | +1 | F |

| | $Q(\text{C}, \textit{left})$ | $Q(\text{C}, \textit{jump})$ | $Q(\text{E}, \textit{left})$ | $Q(\text{E}, \textit{right})$ | $Q(\text{F}, \textit{left})$ | $Q(\text{F}, \textit{right})$ |
|---|---|---|---|---|---|---|
| Initial | 0 | 0 | 0 | 0 | 0 | 0 |
| Transition 1 | | | | | | |
| Transition 2 | | | | | | |
| Transition 3 | | | | | | |
| Transition 4 | | | | | | |

# Logic

## سوال اول) فرزندان محمد

وقتی از محمد درباره سن فرزندانش پرسیدند، او گفت: «آلیس کوچکترین فرزند من است، به شرطی که بیل کوچکترین نباشد. همچنین آلیس کوچکترین فرزند من نیست، اگر کارل کوچکترین نباشد.» دانش پایه‌ای برای توصیف این مسئله و این واقعیت که فقط یکی از این سه فرزند می‌تواند کوچکترین باشد را بنویسید. سپس با استفاده از الگوریتم resolution، نشان دهید که بیل کوچکترین فرزند اوست.

## سوال دوم) مجله معمایی

در انتهای یک مجله معمایی را می‌بینید: «فرض کنید دروغ‌گوها همیشه چیزی را که غلط است می‌گویند و راست‌گوها همیشه حقیقت را می‌گویند. همچنین فرض کنید که امین یا دروغ‌گو است یا راست‌گو.» این معما سپس حقایق دیگری را درباره امین ارائه می‌دهد و می‌پرسد که آیا امین باید راست‌گو باشد؟ شما این حقایق را به منطق گزاره‌ای تبدیل کرده و یک روش حل را بر روی رایانه اجرا می‌کنید. از آنجایی که اشتباهی مرتکب نمی‌شوید، رایانه پاسخ صحیح را به شما می‌دهد. شما از رایانه می‌پرسید که آیا حقایق به این نتیجه می‌رسند که امین راست‌گو است.

الف) رایانه به شما می‌گوید که حقایق به این نتیجه می‌رسند که امین راست‌گو است. از آنجا که متن بیان کرده که امین یا دروغ‌گو است یا راست‌گو، آیا می‌توانید نتیجه بگیرید که امین دروغ‌گو نیست؟

ب) رایانه به شما می‌گوید که حقایق به این نتیجه نمی‌رسند که امین راست‌گو است. از آنجا که متن بیان کرده که امین یا دروغ‌گو است یا راست‌گو، آیا می‌توانید نتیجه بگیرید که امین دروغ‌گو است؟

## سوال سوم) صداهای بوق الکترونیکی

گزاره‌های زیر را در نظر بگیرید که در آن دو گزاره به زبان گفتاری و دو گزاره به صورت منطق مرتبه اول[1] ارائه شده است.

1) همه ربات‌های کت‌بوت (CatBot robots) در شب صداهای بوق الکترونیکی تولید می‌کنند.

2) $\forall x \forall y (Have(x, y) \wedge Real\_Cat(y) \Rightarrow \sim \exists z (Have(x, z) \Rightarrow Mice(z)))$.

3) افراد سبک‌خواب (Light sleepers) هیچ چیزی که در شب صداهای بوق الکترونیکی تولید کند، ندارند.

4) سوزی (Susie) یا یک گربه واقعی (Real Cat) یا یک ربات کت‌بوت (CatBot robot) دارد.

نتیجه-) $Light\_Sleeper(Susie) \Rightarrow \sim \exists z (Have(Susie, z) \wedge Mice(z))$.
5

الف) گزاره‌های 1، 3 و 4 را به صورت فرمول‌های خوش‌ساختار[2] در منطق مرتبه اول با استفاده از گزاره‌های زیر بنویسید:

- CatBot_Robot(x)

- Have(x, y)

- Make_Noise(x)

- Real_Cat(x)

- Light_Sleeper(x)

سپس گزاره 2 و نتیجه-5 را به زبان فارسی بنویسید.

ب) هر فرمول خوش‌ساختار را با معرفی ثابت‌ها به جای کوانتورهای وجودی (اسکولم‌سازی ساده)، و بازنویسی همه گزاره‌ها به صورت CNF تبدیل کنید.

ج) نتیجه را با استفاده از رزولوشن اثبات کنید. در این مرحله، باید پنج گزاره به صورت CNF به عنوان عبارات پایگاه دانش، و سه گزاره CNF به عنوان نتیجه داشته باشید. لطفاً هنگام استفاده از قانون

---

[1] First Order Logic

[2] Well-Formed Formula

رزولوشن در اثبات خود، به شماره گزاره‌ها اشاره کنید و هنگام پیشروی در اثبات خود، گزاره‌های جدیدی که به دست می‌آورید را شماره‌گذاری کنید.